Testing for homogeneity of variance in time series: Long memory, wavelets, and the Nile River

B. Whitcher,[1] S. D. Byers,[2] P. Guttorp,[3] and D. B. Percival[4]

Received 13 March 2001; revised 9 November 2001; accepted 9 November 2001; published 15 May 2002.

[1] We consider the problem of testing for homogeneity of variance in a time series with long memory structure. We demonstrate that a test whose null hypothesis is designed to be white noise can, in fact, be applied, on a scale by scale basis, to the discrete wavelet transform of long memory processes. In particular, we show that evaluating a normalized cumulative sum of squares test statistic using critical levels for the null hypothesis of white noise yields approximately the same null hypothesis rejection rates when applied to the discrete wavelet transform of samples from a fractionally differenced process. The point at which the test statistic, using a nondecimated version of the discrete wavelet transform, achieves its maximum value can be used to estimate the time of the unknown variance change. We apply our proposed test statistic on five time series derived from the historical record of Nile River yearly minimum water levels covering 622–1922 A.D., each series exhibiting various degrees of serial correlation including long memory. In the longest subseries, spanning 622–1284 A.D., the test confirms an inhomogeneity of variance at short time scales and identifies the change point around 720 A.D., which coincides closely with the construction of a new device around 715 A.D. for measuring the Nile River. The test also detects a change in variance for a record of only 36 years. *INDEX TERMS:* 1829 Hydrology: Groundwater hydrology; 1869 Hydrology: Stochastic processes; 3299 Mathematical Geophysics: General or miscellaneous; *KEYWORDS:* wavelets, time series, change point, Nile River, multiresolution, variance

In spite of all the changing, uncertain, and erroneous factors that must be considered in connection with records of stages of the Nile River, it is believed that they disclose some important information; and there is a fair prospect that they may yield more data with further study and the cumulation of ideas of various students.

Jarvis [1936, p. 1028]

1. Introduction

[2] The words of *Jarvis* [1936] are very prophetic, for, in fact, data collected from the Nile River have spurred the development of a whole field of mathematics (i.e., fractional Brownian motion and fractional Gaussian noise) along with a field of statistics concerned with the behavior of long memory time series. Gathered by *Toussoun* [1925], there exists a remarkable hydrological time series of minimum and maximum water levels for the Nile River. Starting in 622 A.D. and ending in 1922 A.D., the first missing observation in the annual minima occurs in 1285 A.D. This leaves us with several continuous records to analyze; the longest one (663 years) is shown in Figure 1.

[3] A reasonable amount of literature has been written about *Toussoun*'s [1925] Nile River data. Some notable facts are provided here. The minimum water levels for the Nile River are not actually the yearly minima. These values were recorded around the end of June each year, whereas the maximum water levels were the actual yearly maxima [*Popper*, 1995] even though *Brooks* [1949, p. 329] notes that the lowest levels of the Nile occur in April and May with erratic behavior in June and the beginning of July. Various time domain and spectral domain methods have been used to analyze these data. Given the current state of knowledge about this process and its apparent long memory structure, these past results will largely be ignored.

[4] A brief history of the beginning portion of *Toussoun*'s [1925] record (622–1284 A.D.) is provided here. Although long-range dependence was initially discovered in the Nile River series by *Hurst* [1951], statistical modeling of this time series was first done as a self-similar process in terms of a fractional Gaussian noise (FGN) model [*Mandelbrot and van Ness*, 1968; *Mandelbrot and Wallis*, 1969] in the doctoral works of *Mohr* [1981] and *Graf* [1983]. This time series model is briefly discussed in section 2. Both Mohr and Graf used the Fourier transform to estimate the self-similarity parameter of an FGN model. *Graf* [1983] reported estimates of $H = d + 1/2$ between 0.83 and 0.85, where H is the so-called Hurst parameter for an FGN model. *Beran* [1994, p. 118] has reported estimates of $H = 0.84$ for FGN and $H = 0.90$ for a realization of a fractionally integrated autoregressive, moving average (ARFIMA) model with 95% confidence intervals of (0.79, 0.89) and (0.84, 0.96), respectively. He also established a goodness-of-fit test for the spectral density function of a long memory process. An approximate p value for the FGN model of the yearly minimum water levels of the Nile River is 0.70, meaning that FGN appears to fit the spectral density function of the Nile River series well.

[5] An important assumption behind any stationary process is that its variance is a constant independent of the time index t. Time series models based on real data may be used to simulate the process and test, for example, hydrological systems. The impact on equipment and the population therefore depends on an accurate

[1]Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, Colorado, USA.
[2]AT&T Labs Research, Florham Park, New Jersey, USA.
[3]Department of Statistics, University of Washington, Seattle, Washington, USA.
[4]Applied Physics Laboratory, University of Washington, Seattle, Washington, USA.

0043-1397/02/2001WR000509$09.00

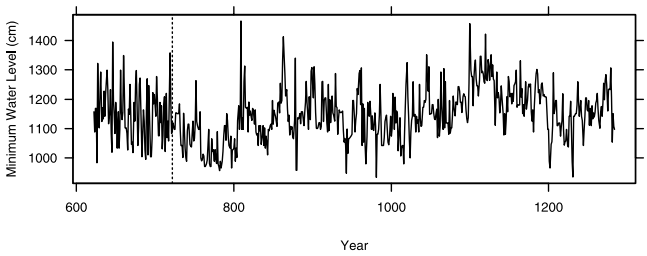

Figure 1. Nile River minimum water levels for 622–1284 A.D. The vertical dotted line is at 730 A.D.

depiction of the true process. The adequacy of any particular time series model may be compromised by nonstationary behavior. From Figure 1 it is apparent that the first 100 or so observations (before the vertical dotted line) exhibit more variability than the remaining 563. Although visually striking, this apparent change in variance occurs in the presence of long-range dependence. In the context of short memory models, such as stationary autoregressive moving average (ARMA) processes [*Box and Jenkins*, 1976], a number of methods have been proposed for testing homogeneity of variance. For a time series consisting of either independent Gaussian random variables with zero mean and possibly time-dependent variances σ_t^2 or a moving average of such variables, *Nuri and Herbst* [1969] proposed to test the hypothesis that σ_t^2 is constant for all t by using the periodogram of the squared random variables. *Wichern et al.* [1976] proposed a moving-block procedure for detecting a single change of variance at an unknown time point in an autoregressive model of order one. *Hsu* [1977, 1979] looked at detecting a single change in variance at an unknown point in time in a series of independent observations.

[6] *Davis* [1979] studied tests for a single change in the innovations variance at a specified point in time in an autoregressive (AR) process. *Abraham and Wei* [1984] used a Bayesian framework to study changes in the innovation variance of an ARMA process. *Tsay* [1988] looked at detecting several types of disturbances in time series, among them variance changes, by analyzing the residuals from fitting an ARMA model. *Inclán and Tiao* [1994] investigated the detection of multiple changes of variance in sequences of independent Gaussian random variables by recursively applying a cumulative sum of squares test to pieces of the original series. Recently, *Perreault et al.* [2000] used Bayesian methods to detect and locate variance changes in sequences of independent Gaussian random variables with application to hydrometeorological time series.

[7] The test we propose, to be discussed in section 4, may be regarded as an adaptation of the work of *Hsu* [1977, 1979] and *Inclán and Tiao* [1994] to handle long memory processes, but, except for the tests with a null hypothesis of white noise with constant variance, it is not an easy matter to adapt the other tests. As noted by *Tang and MacNeill* [1991], serial correlation can produce striking effects in the distributions of change-point statistics. Testing processes with spectral density functions (SDFs) dominated by low-frequency content, like those associated with long memory processes, increases the type 1 error rate, that is, the null hypothesis of no change in the parameter of interest is rejected more often than expected when, in fact, no change has occurred. *Tang and MacNeill* [1991] go on to develop an "adjustment" for their test statistic on the basis of the SDF of a process with absolutely summable autocovariances. Change points in the mean level and serial correlation of annual discharges of the Nile River at Aswan are investigated by *MacNeill et al.* [1991]. Since long memory processes do not have absolutely summable autocovariances, this approach is not applicable here.

[8] A key assumption behind several of the tests proposed previously is accurately fitting a parametric model (e.g., ARMA process) to the observed time series. An incorrect specification of the statistical model may lead to erroneous judgments of variance change points when analyzing the residuals. We avoid this model-fitting step by testing the coefficients from a multiscale decomposition of the original time series. By first applying this decomposition, we may apply a relatively simple testing procedure to a wide range of stochastic processes without committing to a particular time series model. Each level of the decomposition is associated with a specific frequency range in the original process. Although we are not testing observations from this process directly, hypothesis testing results from each level of the decomposition may be directly associated with features in the original process. For example, variance changes in lower levels of the decomposition correspond with higher frequencies in the original process and imply a change of variance. If a variance change was detected at higher levels of the decomposition, this would imply a change in the low-frequency or long memory content of the original process. The multiscale decomposition removes the serial correlation of a long memory process and allows us to test for a variety of departures from stationarity.

[9] Specifically, in this paper we demonstrate how the discrete wavelet transform (DWT) may be used to construct a test for homogeneity of variance in a time series exhibiting long memory characteristics. The DWT is a relatively new tool for time series analysis but has already proven useful for investigating other types of nonstationary events. For example, *Wang* [1995] tested wavelet coefficients at fine scales to detect jumps and sharp cusps of signals embedded in Gaussian white noise, and *Ogden and Parzen* [1996] used wavelet coefficients to develop data-dependent thresholds for removing noise from a signal. The key property of the DWT that makes it useful for studying possible nonstationarities is that it transforms a time series into coefficients that reflect changes at various scales and at particular times. For fractional difference (FD) and related long memory processes the wavelet coefficients for a given scale are approximately uncorrelated (see, for example, *Tewfik and Kim* [1992], *McCoy and Walden* [1996], *Wornell* [1996], and our discussion in section 4.2). We show here that this approximation is good enough that a test designed for a null hypothesis of white noise may be used for testing homogeneity of variance in a long memory process on a scale by scale basis. An additional advantage of testing the output from the DWT is that the scale at which the inhomogeneity occurs may be identified. Using a variation of the DWT, we also investigate an auxiliary test statistic that can estimate the time at which the variance of a time series changes.

[10] While we will concentrate on FD processes in order to validate our proposed homogeneity of variance test, in fact, our test is by no means limited to just these processes. The key to the methodology proposed here is the decorrelation property of the DWT, which can be verified for other processes by using equation (40) and by comparing the results with Table 3. Alternatively, the DWT yields an octave band decomposition of the spectrum, that is, the DWT partitions the spectrum of a process such that the wavelet coefficients are associated with disjoint frequency intervals $f \in [1/2^{j+1}, 1/2^j]$ for $j = 1, 2, \ldots$. We can expect the decorrelation property of the DWT to hold for level j wavelet coefficients as long as there is relatively little change in the spectrum of the process over the frequency interval $[1/2^{j+1}, 1/2^j]$, as is true for FD processes. Other

processes for which the decorrelation property holds include first-order AR processes with nonnegative lag one autocorrelations, FGN, stationary long memory power law processes, and certain ARFIMA processes (i.e., extensions to FD processes that do not exhibit rapid variations within octave bands). Loosely speaking, the spectrum of the process must increase as frequency approaches 0 and must not vary too much within the specified frequency intervals.

[11] An outline of the remainder of this paper is as follows: A review of long memory processes, such as FD processes, FGN, and pure power law processes, is provided in section 2. Section 3 provides a brief introduction to the DWT and points out some key properties we use later on. Section 4 defines the test statistic for detecting sudden changes of variance in a sequence of independent Gaussian random variables, summarizes some simulation results for the DWT of FD processes, and applies our proposed test to five time series from the historical record of Nile River minimum water levels. Section 5 introduces a procedure for determining the location of a variance change, presents simulation results for FD processes with a known variance change, and demonstrates the procedure using the Nile River data. Section 6 gives some concluding remarks.

2. Time Series Models for Long Memory Processes

[12] Long memory structure in a stationary process may be characterized via the autocovariance sequence, or equivalently, its spectral density function. To be more precise, we say that Y_t is a stationary long memory process if there exists a real number $-1 < \alpha < 0$ and a constant $C_S > 0$ such that

$$\lim_{f \to 0} \frac{S_Y(f)}{C_S |f|^{\alpha}} = 1 \tag{1}$$

[*Beran*, 1994, p. 42]. Given the equivalence of the autocovariance sequence and SDF for covariance stationary processes, an alternative formulation of equation (1) is possible in terms of its autocovariance sequence $s_{Y,\tau}$. We say that Y_t is a stationary long memory process if there exists a real number $-1 < \beta < 0$ and a constant $C_s > 0$ such that

$$\lim_{\tau \to \infty} \frac{s_{Y,\tau}}{C_s \tau^{\beta}} = 1 \tag{2}$$

[*Beran*, 1994, pp. 42–43]. These two definitions form a basis for describing long memory structure in observed time series. Low frequencies in the spectrum, corresponding to oscillations with longer and longer periods, exhibit increasing strength inversely proportional to frequency. This corresponds to a slow rate of decay in the autocovariance sequence, and thus observations separated by long time spans still exhibit nonnegligible covariance.

[13] Suppose we have a time series that we are considering to model as a realization of one portion $Y_0, \ldots, Y_{N-1}$ of a Gaussian fractionally integrated autoregressive, moving average (ARFIMA) process Y_t. Typically, there are three parameters used to characterize the ARFIMA process (p, d, and q). The first parameter p indicates how many autoregressive terms are present, the second parameter d determines the degree of integration, and the third parameter q indicates how many moving average terms are included. For our purposes here, assume $p = q = 0$ and that d is allowed to be any positive real number; then the ARFIMA process is the so-called fractionally differenced process. This process may be represented as

$$Y_t = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \epsilon_{t-k}, \tag{3}$$

where $-1/2 < d < 1/2$, and ϵ_t is a Gaussian white noise process with mean zero and variance σ_ϵ^2. The SDF for this process is given by

$$S_Y(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2d}} \quad \text{for} \quad |f| \le \frac{1}{2}, \tag{4}$$

while its autocovariance sequence $s_{Y,\tau}$ may be obtained using the recursion

$$s_{Y,\tau} = s_{Y,\tau-1} \frac{\tau + d - 1}{\tau - d}, \quad \tau = 1, 2, \ldots, \tag{5a}$$

$$s_{Y,0} = \frac{\sigma_\epsilon^2 \Gamma(1-2d)}{\Gamma^2(1-d)} \tag{5b}$$

(for $\tau < 0$ we have $s_{Y,\tau} = s_{Y,-\tau}$). When $0 < d < 1/2$, the SDF has a singularity at 0, in which case the process exhibits slowly decaying autocovariances and constitutes a simple example of a stationary long memory process according to *Beran* [1994] [see also *Granger and Joyeux*, 1980; *Hosking*, 1981]. In this case, d is called the long memory parameter. Figure 2 displays several examples of spectra from FD processes with $0 < d < 1/2$. When plotted on a log-log scale, the SDFs are linear for frequencies $0 < f < 1/4$.

[14] Another popular model for a long memory process is fractional Gaussian noise, introduced by *Mandelbrot and van Ness* [1968] and more recently investigated by *Molz et al.* [1997]. Suppose that $Y_0, \ldots, Y_{N-1}$ is one portion of an FGN; then it is a stationary process with autocovariance sequence given by

$$s_{Y,\tau} = \frac{\sigma_Y^2}{2} \left(|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H} \right), \tag{6}$$

$\tau = 0, \pm 1, \pm 2, \ldots$, where $\sigma_Y^2 > 0$ is the variance of the process and $0 < H < 1$ is the Hurst parameter. The spectrum of an FGN may be derived from an appropriately filtered SDF of fractional Brownian motion and evaluated via numeric integration, although finite-term approximations may be used in practice. When $H = 1/2$, an FGN is equivalent to a white noise process with $s_{Y,\tau} = 0$ for $|\tau| > 0$. When $1/2 < H < 1$, Y_t satisfies equation (2) and constitutes an example of a long memory process. A third example a stationary long memory process is a pure power law (PPL) process and is defined by appealing to equation (1) directly. Its SDF has the form

$$S_Y(f) = C_S |f|^{\alpha}, \quad \text{for} \quad |f| \le \frac{1}{2}, \tag{7}$$

where $C_S > 0$ and is linear for all $|f| \le 1/2$. When $-1 < \alpha < 0$, the PPL process is stationary and obviously obeys equation (1).

[15] All three models presented here are similar in the fact that they are characterized by two parameters: one for the exponent of the power law across frequencies, and the second concerns the variance of the process. Careful choice of both parameters allows one to produce SDFs with almost identical features for all three models. In spite of these similarities, FD processes have several advantages:

1. Both the SDF and (autocovariance sequence) for an FD process are easy to compute.

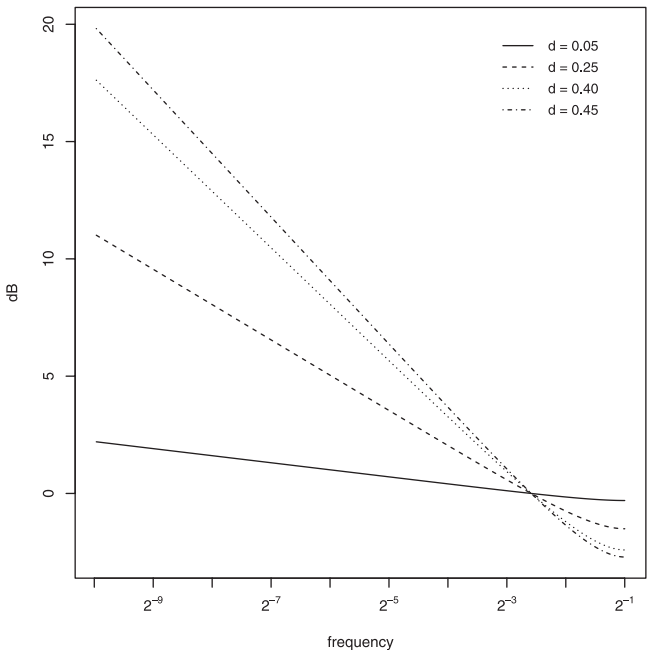

Figure 2. Spectral density functions for fractional difference (FD) processes with long memory parameter $d \in \{0.05, 0.25, 0.4, 0.45\}$. The x axis is displayed on the $\log_2$ scale, and the y axis is in decibels.

2. There is a natural extension to FD processes to cover nonstationary processes with stationary backward differences.

3. An FD process may be considered a special case of an autoregressive, fractionally integrated, moving average process, which involves additional parameters that allow flexible modeling of high-frequency content of a time series.

[16] For these reasons the FD process is the model of choice, being used to describe the ability of the DWT to decorrelate a long memory process (section 3.2) and in all simulation studies provided here.

3. Discrete Wavelet Transforms

[17] The orthonormal discrete wavelet transform has been successfully applied to time series analysis in geophysics because of its ability to separate variability across timescales [see, e.g., *Kumar*, 1996; *Kumar and Foufoula-Georgiou* [1997]; *Whitcher et al.*, 2000a]. The DWT is a way of decomposing a time series into coefficients associated with time and a specific frequency band (usually called a scale). The discrete Fourier transform, on the other hand, decomposes a time series into coefficients associated with a Fourier frequency and no time information. In one sense the discrete Fourier transform concentrates on frequency resolution while the DWT gives up frequency resolution in order to gain some time resolution. We utilize the ability of the DWT to isolate features local in both time and scale in order to capture specific nonstationarities. We are strictly interested in discrete time series and therefore restrict ourselves exclusively to the DWT. Introductions to the continuous wavelet transform and its relation to the DWT are given by, e.g., *Chui* [1997] and *Mallat* [1998].

3.1. The Discrete Wavelet Transform

[18] Let $h_1 = (h_{1,0}, \ldots, h_{1,L-1}, 0 \ldots, 0)^T$ denote the wavelet filter coefficients of a Daubechies compactly supported wavelet for unit scale [*Daubechies*, 1992, chapter 6] with 0 padded to length N by defining $h_{1,l} = 0$ for $l \geq L$. A wavelet filter must satisfy the following three basic properties:

$$\sum_{l=0}^{L-1} h_{1,l} = 0, \tag{8}$$

$$\sum_{l=0}^{L-1} h_{1,l}^2 = 1, \tag{9}$$

and for all nonzero integers n

$$\sum_{l=0}^{L-1} h_{1,l} h_{1,l+2n} = 0. \tag{10}$$

That is, a wavelet filter must sum to 0 (have zero mean), must have unit energy, and must be orthogonal to its even shifts. The most simple example of a wavelet filter h_1 is the Haar wavelet given by

$$h_{1,0} = \frac{1}{\sqrt{2}} \quad \text{and} \quad h_{1,0} = \frac{-1}{\sqrt{2}} \tag{11}$$

and is equivalent to a Daubechies extremal phase wavelet filter of length $L = 2$ (we denote this family of wavelets as "D(L)" for length L).

[19] Let

$$H_{1,k} = \sum_{l=0}^{N-1} h_{1,l} e^{-i2\pi lk/N}, \quad k = 0, ..., N-1, \tag{12}$$

be the discrete Fourier transform of h_1, also known as the transfer or gain function. Let $g_1 = (g_{1,0}, \ldots, g_{1,L-1}, 0, \ldots, 0)^T$ be the zero padded scaling filter coefficients, defined via $g_{1,l} = (-1)^{l+1} h_{1,L-1-l}$ for $l = 0, \ldots, L-1$, and let $G_{1,k}$ denote its gain function. The gain function H_1 of the Haar wavelet (equation (11)) indicates which frequencies are associated with this filter. Figure 3 shows the real-valued squared gain functions $\mathcal{H}_1 = |H_1|^2$ and $\mathcal{G}_1 = |G_1|^2$ for the Haar wavelet and scaling filters, respectively. An ideal high-pass filter would only capture frequencies $1/4 < f < 1/2$, but the Haar wavelet filter is a very poor approximation to this ideal filter. Hence frequencies <1/4 are also given positive weight. When filtering a time series via the DWT and using the Haar wavelet filter, significant amounts of low-frequency content may leak into the resulting wavelet coefficients. To overcome this problem, and to obtain better frequency resolution, *Daubechies* [1992] developed methods to produce compactly supported wavelets of various lengths. For example, the D(4) wavelet is given by

$$\begin{aligned} h_{1,0} &= \frac{1-\sqrt{3}}{4\sqrt{2}}, \quad h_{1,1} = \frac{-3+\sqrt{3}}{4\sqrt{2}}, \\ h_{1,2} &= \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad \text{and} \quad h_{1,3} = \frac{-1-\sqrt{3}}{4\sqrt{2}}. \end{aligned} \tag{13}$$

Figure 3 shows the squared gain functions for the D(4) wavelet and scaling filters. Their approximation to ideal band-pass filters is improved over the Haar wavelet and scaling filters. Further improvements in frequency resolution may be found by using even longer wavelets.

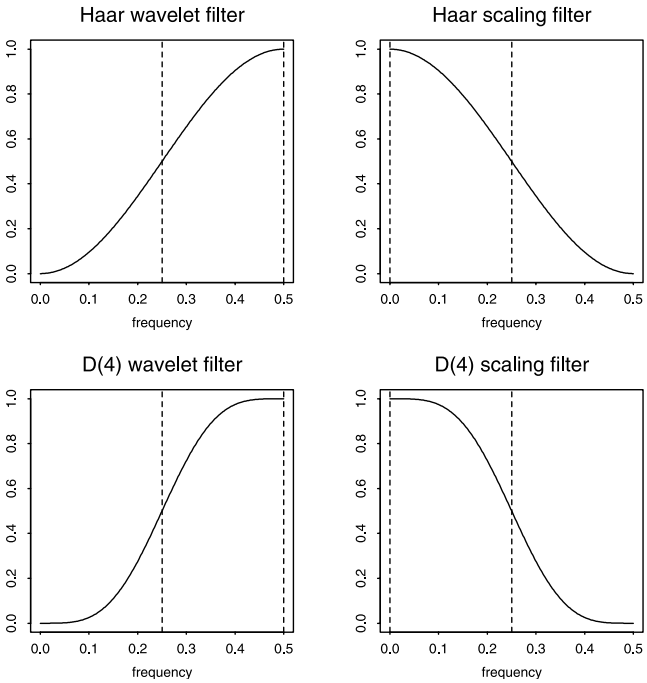

Figure 3. Squared gain functions for the wavelet and scaling filters from the Haar and D(4) families of compactly supported wavelets. The dotted lines represent the frequency interval associated with an ideal high-pass filter for the wavelet filter and an ideal low-pass filter for the scaling filter.

[20] Working from the unit-scale wavelet filter h_1, one may obtain a wavelet filter associated with higher scales (lower frequencies) using a simple recursion in the frequency domain. Define the length N wavelet filter h_j for scale $\tau_j = 2^{j-1}$ as the inverse DFT of

$$H_{j,k} = H_{1,2^{j-1}k \bmod N} \prod_{l=0}^{j-2} G_{1,2^l k \bmod N}, \tag{14}$$

$k = 0, \ldots, N-1$, where "k mod N" stands for "k modulo N" and forces the assumption of periodicity on the gain functions. When $N > L_j = (2^j - 1)(L-1) + 1$, the last $N - L_j$ elements of h_j are zero, so the wavelet filter h_j has at most L_j nonzero elements. At each scale j the wavelet filter h_j is capturing smaller and smaller portions of low-frequency content from the original signal and thus approximates an ideal band-pass filter with passband $f \in [1/2^{j+1}, 1/2^j]$. Also, let us define the scaling filter g_J for scale τ_J as the inverse DFT of

$$G_{J,k} = \prod_{l=0}^{J-1} G_{1,2^l k \bmod N}, \quad k = 0, \ldots, N-1. \tag{15}$$

After applying a succession of wavelet filters for scales $j = 1, \ldots, J$, the scaling filter captures the remaining low-frequency content from the signal and approximates an ideal low-pass filter with passband $f \in (0, 2^{j+1})$. Whereas wavelet coefficients associated with band-pass filters will have mean zero, under appropriate assumptions of the underlying spectrum of the process the scaling coefficients will contain the mean of the process.

[21] Let $Y_0, \ldots, Y_{N-1}$ be a time series of length N. For scales such that $N \geq L_j$ we can filter the time series using h_j to obtain the wavelet coefficients

$$W_{j,t} = 2^{j/2} \widetilde{W}_{j,2^j(t+1)-1}$$

and

$$\left\lceil (L-2)\left(1 - \frac{1}{2^j}\right) \right\rceil \leq t \leq \left\lfloor \frac{N}{2^j} - 1 \right\rfloor, \tag{16}$$

where

$$\widetilde{W}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j - 1} h_{j,l} Y_{t-l}, \quad t = L_j - 1, \ldots, N-1. \tag{17}$$

The $W_{j,t}$ coefficients are obtained by downsampling every 2^jth value of the $\widetilde{W}_{j,t}$ coefficients, which form a portion of one version of a "nondecimated" DWT called the "maximal overlap" DWT (see *Percival and Guttorp* [1994] and *Percival and Mofjeld* [1997] for more details on this transform). Other versions of the nondecimated DWT are discussed by *Shensa* [1992], *Coifman and Donoho* [1995], *Nason and Silverman* [1995], and *Bruce and Gao* [1996].

[22] The $W_{j,t}$ coefficients derived from a Daubechies family of wavelets are associated with generalized differences (changes) on a scale of length τ_j. For example, the unit-scale Haar DWT coefficients $W_{1,t}$ are differences of adjacent observations in the original series

$$W_{1,t} = (Y_{2t+1} - Y_{2t})/\sqrt{2}, \quad t = 0, 1, \ldots, N/2 - 1, \tag{18}$$

and hence may be thought of as changes at a scale of length $\tau_1 = 2^{1-1} = 1$. The Haar DWT coefficients at the second scale are computed via

$$W_{1,t} = [(Y_{4t+3} + Y_{4t+2}) - (Y_{4t+1} + Y_{4t})]/2, \tag{19}$$

$t = 0,1, \ldots, N/4 - 1$, and correspond to differences of adjacent averages at a scale of length $\tau_2 = 2^{2-1} = 2$. This analogy proceeds for all positive levels j of the DWT that correspond to changes at scales of lengths $\tau_j = 2^{j-1}$. Thus when referring to a particular set of DWT coefficients, the terms "level j" and "scale τ_j" are equivalent. The exception is the scaling coefficients from the final level of the DWT $V_{J,t}$ coming from the application of g_J to Y_t. They capture the low-frequency content of the process and are associated with weighted averages at a scale of length $2\tau_J$.

[23] In practice, the DWT is implemented via a pyramid algorithm [*Mallat*, 1989] illustrated in Figure 4. First, the data vector $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_{N-1})^T$ is convolved with the filters h_1 and g_1, whose gain functions are given by equation (14) and denoted by H_1 and G_1, respectively. The resulting filtered series are reduced to half their original lengths via downsampling by two. The downsampled output from the h_1 filter is the length $N/2$ vector of wavelet coefficients $\mathbf{W}_1$, and the downsampled output from the g_1 filter is the length $N/2$ vector of scaling coefficients $\mathbf{V}_1$. The vector of scaling coefficients $\mathbf{V}_1$ is filtered and downsampled again, using h_1 and g_1, to yield the length $N/4$ vectors $\mathbf{W}_2$ and $\mathbf{V}_2$, respectively. The scaling coefficients are recursively filtered and downsampled until only one wavelet coefficient W_J and one scaling coefficient V_J remain. All wavelet coefficient vectors and

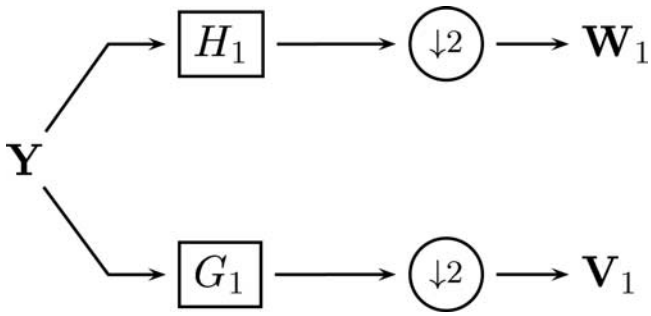

Figure 4. Flow diagram illustrating the decomposition of a discrete time series $\mathbf{Y}$ into the unit-scale wavelet coefficients $\mathbf{W}_1$ and the unit-scale scaling coefficients $\mathbf{V}_1$. The time series $\mathbf{Y}$ is filtered using the wavelet filter with gain function H_1 and every other value removed (downsampled by 2) to produce the length $N/2$ wavelet coefficient vector $\mathbf{W}_1$. Similarly, $\mathbf{Y}$ is filtered using the scaling filter with gain function G_1 and is downsampled to produce the length $N/2$ vector of scaling coefficients $\mathbf{V}_1$.

the final scaling coefficient are kept and may be organized into the length N vector $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{J-1}, \mathbf{W}_J, \mathbf{V}_J)^T$, where $\mathbf{W}_j$ is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\tau_j = 2^{j-1}$. A simple modification to the pyramid algorithm for computing DWT coefficients, namely, not downsampling the output at each scale and inserting zeros between coefficients in h_1 and g_1, yields the algorithm for computing $\widetilde{W}_{j,t}$, as described by *Percival and Mofjeld* [1997].

[24] An additive decomposition of $Y_0, \ldots, Y_{N-1}$ may be obtained by performing a multiresolution analysis (MRA) on the observed process. We concentrate on an MRA based on the nondecimated DWT since the particular shape of the wavelet filter used does not heavily influence the results. Once the nondecimated DWT has been performed, simply convolve the $\widetilde{W}_{j,t}$ coefficients with h_j to produce the wavelet detail

$$\widetilde{\mathcal{D}}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{N-1} h_{j,l} \widetilde{W}_{j,t+l \bmod N}, \tag{20}$$

where $t = 0,1, \ldots, N-1$. Convolving the $\widetilde{V}_{J,t}$ coefficients with g_J produces the wavelet smooth

$$\widetilde{\mathcal{S}}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{N-1} g_{J,l} \widetilde{V}_{J,t+l \bmod N}, \tag{21}$$

where $t = 0,1, \ldots, N-1$. The subseries $\widetilde{\mathcal{D}}_j$ and $\widetilde{\mathcal{S}}_J$ in an MRA form an additive decomposition of the original time series via

$$Y_t = \sum_{j=1}^{J} \widetilde{\mathcal{D}}_{j,t} + \widetilde{\mathcal{S}}_{J,t}. \tag{22}$$

Each subseries $\widetilde{\mathcal{D}}_j$ is associated with changes at scale $\tau_j = 2^{j-1}$, while $\widetilde{\mathcal{S}}_J$ is associated with weighted averages over scales of 2^J (see *Percival and Mofjeld* [1997] for more details).

3.2. The DWT of Long Memory Processes

[25] The ability of the wavelet transform to decorrelate time series, such as long memory processes, producing DWT coefficients for a given scale that are approximately uncorrelated is well known [see, e.g., *Tewfik and Kim*, 1992; *McCoy and Walden*, 1996; *Wornell*, 1996]. Here we explore the output of the DWT when applied to an FD process in terms of the SDF of the DWT coefficients.

[26] Let Y_t be an FD process with SDF given by equation (4) and $0 < d < 1/2$. We know that the DWT coefficients $W_{1,0}, \ldots, W_{1,N/2}$ for unit scale are simply the original time series convolved with the wavelet filter h_1 and downsampled by 2. The SDF of a downsampled process, say $U_t = Y_{2t}$, is given by

$$S_U(f) = \frac{1}{2}\left[S_Y\left(\frac{f}{2}\right) + S_Y\left(\frac{f+1}{2}\right)\right]. \tag{23}$$

We also know from basic Fourier theory that filtering a time series corresponds to a multiplication of its spectrum with the squared gain function of that filter. Hence the filtered coefficients $h_1 * Y_t$ (where the asterisk denotes the convolution operator) have SDF given by $\mathcal{H}_1(f)S_Y(f)$, and the spectrum of the unit-scale DWT coefficients $W_{1,0}, \ldots, W_{1,N/2}$ is therefore

$$S_{Y,1}(f) = \frac{1}{2}\left[\mathcal{H}_1\left(\frac{f}{2}\right)S_Y\left(\frac{f}{2}\right) + \mathcal{H}_1\left(\frac{f+1}{2}\right)S_Y\left(\frac{f+1}{2}\right)\right]. \tag{24}$$

From equation (24) the spectrum of $W_{1,t}$ is the average of two stretched pieces of the filtered spectrum. This averaging is important because it causes $S_{Y,1}(f)$ to be much more flat than the filtered spectrum.

[27] For the class of Daubechies compactly supported wavelets a closed form expression exists for $\mathcal{H}_1(f)$, namely,

$$\begin{aligned}\mathcal{H}_1^{(\mathrm{D})}(f) &= 2\sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f) \\ &= \mathcal{D}^{\frac{L}{2}}(f)\mathcal{C}(f)\end{aligned} \tag{25}$$

[*Daubechies*, 1992, chapter 6.1], where $\mathcal{D}(f) = |2\sin(\pi f)|^2$ is the squared gain function of a first order backward difference filter and

$$\mathcal{C}(f) = \frac{1}{2^{L-1}} \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f). \tag{26}$$

Substituting equations (4) and (25) into equation (24) gives us

$$\begin{aligned}S_{Y,1}^{(\mathrm{D})}(f) = \frac{1}{2}\Bigg[&\mathcal{D}^{\frac{L}{2}}\left(\frac{f}{2}\right)\mathcal{C}\left(\frac{f}{2}\right)S_Y\left(\frac{f}{2}\right) \\ &+\mathcal{D}^{\frac{L}{2}}\left(\frac{f+1}{2}\right)\mathcal{C}\left(\frac{f+1}{2}\right)S_Y\left(\frac{f+1}{2}\right)\Bigg]\end{aligned} \tag{27}$$

$$\begin{aligned}= \frac{1}{2}\Bigg[&\mathcal{D}^{-\left(d-\frac{L}{2}\right)}\left(\frac{f}{2}\right)\mathcal{C}\left(\frac{f}{2}\right) \\ &+\mathcal{D}^{-\left(d-\frac{L}{2}\right)}\left(\frac{f+1}{2}\right)\mathcal{C}\left(\frac{f+1}{2}\right)\Bigg].\end{aligned} \tag{28}$$

The squared gain function $\mathcal{D}^{-\left(d-\frac{L}{2}\right)}(f)$ corresponds to an FD process with long memory parameter $d - \frac{L}{2}$ and $\mathcal{C}(f)$ is the squared gain function of a filter with compact support. Figure 5 shows the SDFs for the filtered processes $h_1 * Y_t$ and the unit-scale DWT coefficients when Y_t is an FD process with $d = 0.4$. Although the spectrum of the FD process rapidly approaches infinity as $f \to 0$, the unit-scale wavelet filter removes most low-frequency content $f < 1/4$. The downsampling inherent in the DWT then averages the

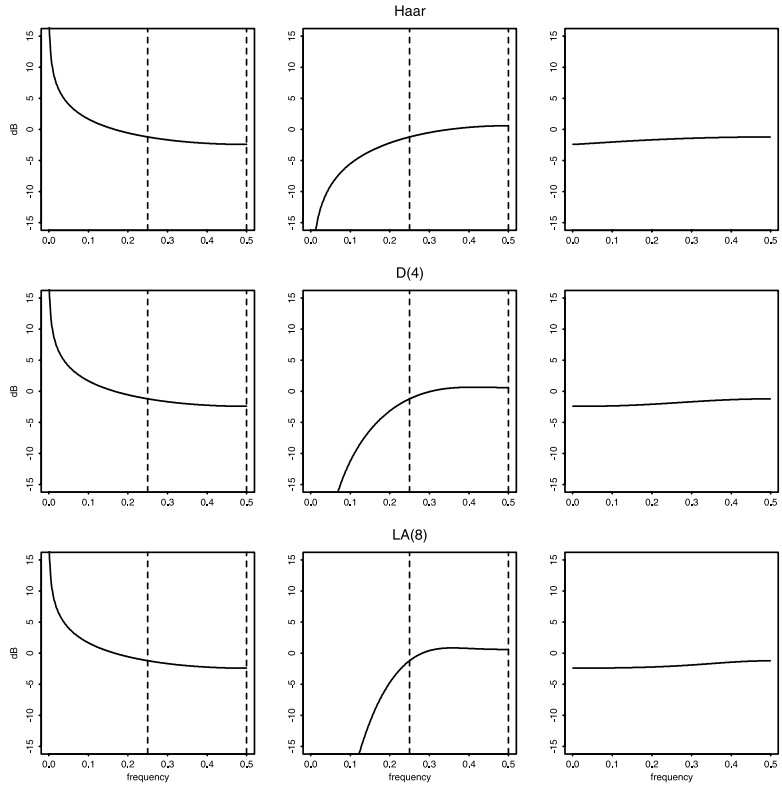

Figure 5. Spectral density functions for (left panels) an FD process Y_t with $d = 0.4$, (center panels) the filtered process $h_1 * Y_t$ using the Haar, D(4), and LA(8) wavelet filters, and (right panels) unit-scale wavelet process $W_{1,t}$ using the Haar, D(4), and LA(8) wavelet filters. The y axis for all plots is plotted on a decibel scale.

two halves of the filtered SDF and results in a spectrum that is difficult to distinguish from a white noise model (i.e., constant).

[28] As an example, consider the spectrum of DWT coefficients at unit scale for an FD process analyzed using the Haar wavelet. It is relatively simple to calculate since $\mathcal{H}_1^{\text{Haar}}(f) = 2\sin^2(\pi f)$. The SDF of the filtered FD process $h_1 * Y_t$ is given by

$$\mathcal{H}_1^{\text{Haar}}(f)S(f) = \frac{1}{2}|2\sin(\pi f)|^{-2(d-1)}. \tag{29}$$

That is, the SDF of the filtered process is proportional to an FD process with parameter $d' = d-1$. Since we were looking at so-called "red" processes (where the spectral energy is dominated by low frequencies) with $0 < d < 1/2$, this means $-1 < d' < -1/2$, and hence the filtered process is "blue" (with spectral energy dominated by high frequencies). The spectrum of the DWT coefficients using the Haar wavelet is therefore

$$S_{Y,1}^{\text{Haar}}(f) = \frac{1}{4}\left[\left|2\sin\left(\frac{\pi f}{2}\right)\right|^{-2(d-1)} + \left|2\cos\left(\frac{\pi f}{2}\right)\right|^{-2(d-1)}\right]. \tag{30}$$

Figure 6 gives the SDFs for the unit-scale Haar DWT coefficients applied to FD processes with $0 \le d \le 0.45$. When $d = 0$, Y_t is a white noise process, and the SDF of the Haar DWT coefficients for unit scale does not depend on f and therefore is also a white noise process. As d increases, the spectrum of the $W_{j,t}$ coefficients deviates from that of white noise, with energy deficient in low-frequency content and slightly dominated by the higher frequencies, but the total variation in the SDF is relatively low (between 1 and 2 dB when $d = 0.45$).

[29] The formula given in equation (24) may be extended to an arbitrary scale τ_j using a simple recursive relation (see Appendix A).

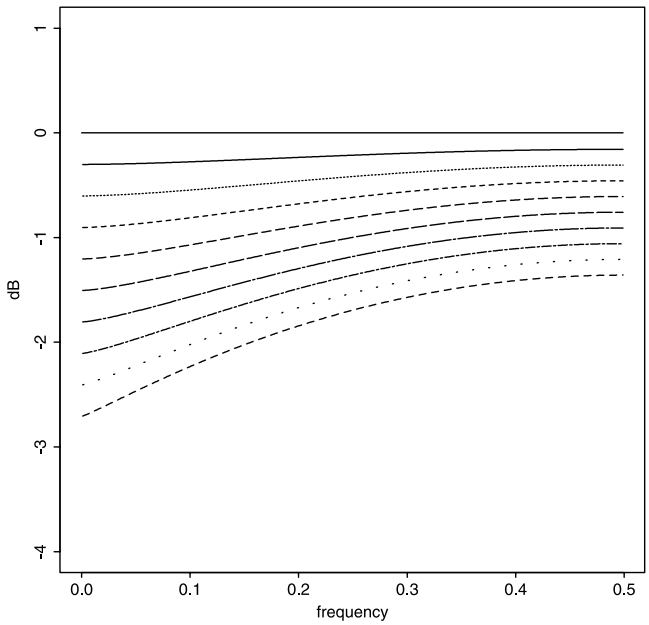

Figure 6. Spectral density functions (in decibels) for wavelet processes $W_{j,t}$ using the Haar wavelet filter from FD processes with, from top to bottom, $d \in \{0, 0.05, 0.1, \ldots, 0.45\}$.

Table 1. Maximum Dynamic Range for the SDF of DWT Coefficients When Applied to FD Processes with Long Memory Parameter $d \in [0, 1/2]$[a]

Level	Haar	D(4)	LA(8)
1	1.48	1.56	1.59
2	2.26	2.56	2.71
3	2.71	2.93	3.00
4	2.87	3.07	3.14
5	2.93	3.10	3.16
6	2.95	3.10	3.16

[a] Values given in decibels.

The SDF for the DWT coefficients $W_{j,0}, \ldots, W_{j,N/2^j}$ of an FD process Y_t, associated with scale τ_j, is given by

$$S_{Y,j}^{(\mathrm{D})}(f) = \frac{1}{2^j} \sum_{k=0}^{2^j-1} \mathcal{H}_j^{(\mathrm{D})}\left(\frac{f+k}{2^j}\right) S_Y\left(\frac{f+k}{2^j}\right), \tag{31}$$

where

$$\mathcal{H}_j^{(\mathrm{D})}(f) = \mathcal{D}^{\frac{L}{2}}(f)\left[\prod_{k=0}^{j-2} 4\cos^2\left(2^k \pi f\right)\right]^{\frac{L}{2}} \times \mathcal{C}\left(2^{j-1} f\right) \mathcal{G}_{j-1}^{(\mathrm{D})}(f). \tag{32}$$

That is, the SDF of the filtered process is stretched by 2^j, and then $2^j - 1$ aliased versions are added to it. This can intuitively be seen through successive applications of equation (23) to the filtered spectrum.

[30] To expand the information contained in Figure 5 to all possible values of $d \in [0,1/2)$ and across wavelet scales, a useful measure to introduce is the dynamic range of a spectrum, defined to be

$$10 \times \log_{10}\left(\frac{\max_f S(f)}{\min_f S(f)}\right). \tag{33}$$

Table 1 gives the maximum dynamic ranges, in decibels, for the spectra of DWT coefficients applied to FD processes with long memory parameter $0 < d < 1/2$. As the level of the DWT increases, the wavelet filter is capturing smaller and smaller ranges of low-frequency content. Even though the SDF of a stationary FD process with $d \in [0,1/2)$ goes to infinity as $f \to 0$, the wavelet filter adapts to these rapid changes so that the dynamic range of the SDF of $W_{j,t}$ is negligible and appears to level off around 3 dB regardless of the wavelet filter used. This lack of dynamic range, which corresponds to almost uncorrelated observations in the wavelet process, is utilized in the next section to test for nonstationary events in the presence of long memory structure.

4. Testing for Homogeneity of Variance

[31] If $Y_0, \ldots, Y_{N-1}$ constitutes a portion of an FD process with long memory parameter $0 < d < 1/2$, and with possibly nonzero mean, then each sequence of wavelet coefficients $W_{j,t}$ for Y_t is approximately a sample from a zero mean white noise process. This enables us to formulate our test for homogeneity of variance using wavelet coefficients for FD processes as follows.

4.1. The Test Statistic

[32] Let $X_0, \ldots, X_{N-1}$ be a time series that can be regarded as a sequence of independent Gaussian (normal) random variables with zero means and variances $\sigma_0^2, \ldots, \sigma_{N-1}^2$. We would like to test the hypothesis

$$H_0 : \sigma_0^2 = \cdots = \sigma_{N-1}^2. \tag{34}$$

A test statistic that can discriminate between equation (34) and a variety of alternative hypotheses (such as H_1: $\sigma_0^2 = \cdots = \sigma_k^2 \neq \sigma_{k+1}^2 = \cdots = \sigma_{N-1}^2$, where k is an unknown change point) is the normalized cumulative sums of squares test statistic D, which has previously been investigated by, among others, *Brown et al.* [1975], *Hsu* [1977], and *Inclán and Tiao* [1994]. Let

$$\mathcal{P}_k = \frac{\sum_{j=0}^{k} X_j^2}{\sum_{j=0}^{N-1} X_j^2}, \quad k = 1, \ldots, N-1, \tag{35}$$

denote the normalized partial energy sequence of X_t. The desired test statistic is given by $D = \max(D^+, D^-)$, where

$$D^+ = \max_{0 \le k \le N-2}\left(\frac{k}{N-1} - \mathcal{P}_k\right) \tag{36}$$

$$D^- = \max_{0 \le k \le N-2}\left(\mathcal{P}_k - \frac{k-1}{N-1}\right). \tag{37}$$

[33] The statistic $\mathcal{P}_k$, as formulated in equation (35), is measuring the accumulation of variance in the time series as a function of time. This quantity is then compared with a line at a 45° angle, where the maximum vertical deviation from this line is recorded as our test statistic D. The rationale is that if the observed time series satisfies H_0, then each random variable should contribute the same amount of variance, and thus $\mathcal{P}_k$ will closely follow the 45° line. If one does not adhere to H_0, for example, if the beginning of the time series exhibits a greater variance than the remainder of the series, then the first few random variables will contribute too much variance to $\mathcal{P}_k$ and deviate drastically from the 45° line. The threshold (critical levels) with which to measure the magnitude of this deviation from the 45° is obtained from standard statistical theory, a brief summary of which is provided below.

[34] For $N > 2$, there is no known tractable closed form expression for the critical levels of D under the null hypothesis. *Brown et al.* [1975] obtained critical levels by an interpolation scheme that makes use of the fact that if N is even and if we group the squared deviates by pairs, then D reduces to the well-known cumulative periodogram test for white noise [*Bartlett*, 1955], for which critical levels are available [*Stephens*, 1974]. *Hsu* [1977] used two methods, Edgeworth expansions and fitting the first three moments to a beta distribution, in order to obtain small sample critical levels for a statistic equivalent to D. Let $\mathbf{P}\{X \le x\}$ denote the probability that the random variable X is less than or equal to a constant x. The results of *Inclán and Tiao* [1994] indicate that, for large N and $x > 0$,

$$\mathrm{P}\left\{\sqrt{N/2} D \le x\right\} \approx \mathbf{P}\left\{\sup_t \left|B_t^0\right| \le x\right\} = 1 + 2\sum_{l=1}^{\infty} (-1)^l e^{-2l^2x^2}, \tag{38}$$

Table 2. Monte Carlo Critical Values With Standard Errors for the Test Statistic $\sqrt{N/2}D$ Using the Haar Wavelet Filter for a Level-α Test[a]

	Sample Size								
α	8	16	32	64	128	256	512	1024	∞
0.10	1.109	1.135	1.157	1.182	1.193	1.197	1.206	1.209	1.224
SE[b]	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	–
0.05	1.232	1.265	1.293	1.313	1.326	1.329	1.345	1.341	1.358
SE	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	–
0.01	1.459	1.508	1.553	1.584	1.596	1.596	1.630	1.617	1.628
SE	0.007	0.008	0.008	0.009	0.008	0.010	0.008	0.007	–

[a] Values are based upon 10,000 replicates.
[b] SE denotes standard errors.

where B_t^0 is a Brownian bridge process, and the right-hand expression is from *Billingsley* [1968, equation (11.39)].

[35] Critical levels for D under the null hypothesis can be readily obtained through Monte Carlo simulations. The procedure presented here is to generate empirical critical values for the cumulative sum of squares test statistic D applied to the first scale of wavelet coefficients from the Haar wavelet filter. The procedure is as follows:

1. Generate an N-length realization of a Gaussian white noise process.
2. Compute the DWT down to scale J using the Haar wavelet filter.
3. Discard all wavelet coefficients at each scale that make explicit use of the periodic boundary conditions.
4. Compute the test statistic D for all scales on the basis of the remaining wavelet coefficients.
5. Calculate the value of D such that 100% $(1-\alpha)$ simulated values are greater than that value. This is the $(1-\alpha)$th quantile.

[36] Table 2 provides Monte Carlo critical levels for comparison with levels determined by a Brownian bridge process. The standard error (SE) is provided for each estimate and was computed via

$$\mathrm{SE} = \sqrt{\frac{\alpha(1-\alpha)}{10,000 f^2}}, \tag{39}$$

where f is the histogram estimate of the probability density function (PDF) at the $(1-\alpha)$th quantile using a bandwidth of 0.01 [*Inclán and Tiao*, 1994]. That is, the variability of the empirical PDF at a specific quantile is estimated by looking at all values from the simulation within a small interval around that quantile. Quantiles of a Brownian bridge process are given at the far right. Since equation (37) is an asymptotic statement, it is of interest to compare critical levels on the basis of Monte Carlo simulations of small sample sizes with these asymptotic critical levels. This provides insight into how quickly the Monte Carlo critical values converge to their asymptotic values. As we can see, there appears to be a slow but steady convergence as the sample size increases.

4.2. Simulation Study Using the DWT

[37] Here we investigate whether, in fact, the DWT of an FD process is a good approximation to white noise as far as performance of the test statistic D is concerned. We first note that the autocovariance between two wavelet coefficients in the same scale is

$$\mathrm{Cov}\{W_{j,t}, W_{j,t+\tau}\} = \sum_{m=-(L_j-1)}^{L_j-1} s_{Y,2^j\tau+m} \times \sum_{l=0}^{L_j-|m|-1} h_{j,l}h_{j,l+|m|}, \tag{40}$$

where $s_{Y,\tau}$ was previously defined in equation (5a). Appendix B provides a derivation for equation (40). Using equation (40), we may compute the unit lag correlations

$$\mathrm{Corr}\{W_{j,t}, W_{j,t+1}\} = \frac{\mathrm{Cov}\{W_{j,t}, W_{j,t+1}\}}{\mathrm{Var}\{W_{j,t}\}} \tag{41}$$

given in Table 3 for the Haar, D(4), and LA(8) wavelet filters and scales 1, 2, 4, and 8 (here "D(4)" and "LA(8)" refer to the Daubechies extremal phase filter with 4 nonzero coefficients and to her least-asymmetric filter with 8 coefficients, respectively) [*Daubechies*, 1992]. Note that all unit lag correlations are negative, with departures from 0 increasing somewhat as j increases.

[38] Computations indicate that $|\mathrm{Corr}\{W_{j,t}, W_{j,t+2}\}| < 0.033$ and $|\mathrm{Corr}\{W_{j,t}, W_{j,t+3}\}| < 0.009$ for all three wavelet filters, and $\mathrm{Corr}\{W_{j,t}, W_{j,t+\tau}\}$ is negligible for $\tau \geq 4$. To ascertain the effect of these small remaining correlations on D, we determined the upper 10%, 5%, and 1% quantiles for the distribution of D on the basis of a large number of realizations of white noise for sample sizes commensurate with time series of length $N \in \{128, 256, 512, 1024, 2048, 2^{15}\}$. Using these quantiles, we then generated a large number of realizations of length N from FD processes with $d \in \{0.05, 0.25, 0.4, 0.45\}$; computed wavelet coefficients for scales 1, 2, 4, and 8 using the three wavelet filters; computed the test statistic D for all four scales on the basis of the wavelet coefficients; and compared the resulting D's to the white noise critical levels (full details about this Monte Carlo study are given by *Whitcher* [1998]). We found the deviations between the actual rejection rates and the rates established for white noise to be generally ~10%, with the agreement decreasing somewhat with increasing scale (this is consistent with Table 3).

[39] Table 4 shows a small portion of the simulation study by *Whitcher* [1998], providing rejection rates when testing for homogeneity of variance on a stationary FD(0.4) process with no variance change. The test statistic D was applied to raw time series of various lengths to correspond with the sample sizes from applying the DWT to an FD(0.4) process of length 128. The rejection rates from the wavelet-based procedure closely follow the nominal rejection rate $\alpha = 0.05$ when using the Monte Carlo critical values from Table 2 and are more conservative than $\alpha = 0.05$ if asymptotic critical values are used. The effect of serial correlation from short memory (ARMA) time series models on test statistics similar to D is also an increase in the type 1 error or probability of false positives [*Johnson and Bagshaw*, 1974; *Tang and MacNeill*, 1991]. By testing the output from the DWT, we are protecting against any time series model with positive autocorrelations at small to large lags.

[40] From these simulation studies we may conduct an approximate α-level test for variance homogeneity of an FD process, on a scale by scale basis, by simply using critical levels determined

Table 3. Lag-One Autocorrelations for Wavelet Coefficients of Scales 1, 2, 4, and 8 for an FD Process With $d = 0.45$ Using the Haar, D(4), and LA(8) Wavelet Filters

Scale	Haar	D(4)	LA(8)
1	−0.0626	−0.0797	−0.0767
2	−0.0947	−0.1320	−0.1356
4	−0.1133	−0.1511	−0.1501
8	−0.1211	−0.1559	−0.1535

Table 4. Comparison of Rejection Rates When Testing for Homogeneity of Variance Using a Stationary FD(0.4) Process[a]

		Monte Carlo			Asymptotic		
N	Raw	Haar	D(4)	LA(8)	Haar	D(4)	LA(8)
64	0.261	0.065	0.044	0.054	0.035	0.031	0.049
32	0.140	0.063	0.051	0.047	0.029	0.047	0.040
16	0.061	0.052	0.053	0.036	0.029	0.025	0.025
8	0.031	0.036	0.045	–	0.016	0.007	–

[a] The test statistic D is applied to raw time series of length N, and the first four levels of the DWT is applied for an FD(0.4) process of length 128 (using both Monte Carlo and asymptotic critical values). All tests are performed at the $\alpha = 0.05$ level of significance. At the smallest sample size the output from an LA(8) DWT produces only two coefficients unaffected by the boundary, and thus the test statistic was not computed.

under the assumption of white noise. No significant difference was detected between the unique wavelet filters used [*Whitcher*, 1998]. Thus the shorter wavelet filters (e.g., Haar or D(2)) are just as effective as the longer filters even though they do not have the same frequency resolution. This is important since longer wavelet filters, although more appropriate for nonstationary processes with stationary differences because the extra implicit differencing operations ensure that the wavelet coefficients form a stationary process with zero mean, produce more boundary coefficients that must be discarded when constructing D via equations (35) and (36), thereby reducing the power of the statistical test.

[41] One may not want to perform Monte Carlo studies in order to obtain critical values for the test statistic D. The simulation study described above was run again substituting the asymptotic critical values (Table 1) for the Monte Carlo critical values. For sample sizes >128 the percentage of times D exceeded the asymptotic critical levels was within 10% of the theoretical quantile. The Haar wavelet filter was found to be conservative for all sample sizes, that is, the percentage of times D exceeded the asymptotic critical levels was below the nominal percentile. Hence wavelet coefficients of length 128 or greater using asymptotic critical values will give reasonable results if Monte Carlo critical values have not been computed.

[42] To investigate how well this approximation performs for large sample sizes, a similar procedure to the one outlined above was performed for FD processes of length $N = 2^{15} = 32{,}768$ using the $W_{j,t}$ coefficients for $j = 1, \ldots, 8$. Owing to the computational time involved, the procedure was only repeated 1000 times. The percentages of times that D exceeded the asymptotic white noise critical levels under these conditions were found to be quite close to the nominal rejection rates with increased variability due to the reduced number of iterations in the Monte Carlo study. Thus all of the simulation studies we have conducted to date indicate that the DWT adequately decorrelates long memory processes for the purpose of using the test statistic D, even for sample sizes as large as 2^{15}.

4.3. Application to the Nile River Water Levels

[43] As an example of a time series exhibiting both long memory and possible inhomogeneity in variance, we consider the Nile River minimum water level time series [*Toussoun*, 1925]. For an interesting recent analysis of this series we direct the reader to *Eltahir and Wang* [1999]. The full historical record from Toussoun spans 622–1925 A.D. with several gaps. There are 5 portions that exceed 30 years in length. The longest consists of $N = 663$ yearly values from 622 to 1284 A.D. and is plotted in Figure 1. *Beran* [1994, p. 118] found an FD model to fit this time series well and obtained an estimate of $d = 0.40$ using an approximate maximum likelihood approach.

[44] Figure 7 shows a wavelet-based MRA of Nile River minimum water levels via equation (22). Each wavelet detail $\widetilde{\mathcal{D}}_j$ is associated with changes at scale $\tau_j = 2^{j-1}$, while $\widetilde{\mathcal{S}}_J$ is associated with weighted averages over scales of $2\tau^J$. We used the D(4) wavelet in conjunction with the nondecimated DWT, extended to N coefficients at each scale by assuming periodic boundary conditions. Visually, it appears that there is greater variability in changes on scales of 1 and 2 years prior to 722 A.D., but not on longer scales. *Beran* [1994, section 10.3] investigated the question of a change in the long memory parameter in this time series by partitioning the first 600 observations into two subseries containing the first 100 and the remaining 500 measurements, respectively. Maximum likelihood estimates of the long memory parameter d were quite different between the two subseries, 0.04 and 0.38, respectively. This analysis suggests a change in d, a conclusion also drawn by *Beran and Terrin* [1996] using a procedure designed to test for change in the long memory parameter.

[45] We can perform a similar analysis using the wavelet variance, which decomposes the variance of Y_t on a scale by scale basis. The wavelet variance ν_j^2 for scale $\tau_j = 2^{j-1}$ is defined to be the variance of $\widetilde{W}_{j,t}$ and may be estimated by

$$\widetilde{\nu}_j^{\,2} = \frac{1}{N - L_j + 1} \sum_{l=L_j-1}^{N-1} \widetilde{W}_{j,l}^2; \tag{42}$$

see *Percival* [1995] for a discussion of the statistical properties of this estimator. The estimated wavelet variances and approximate confidence intervals (derived in Appendix C), given a partitioning

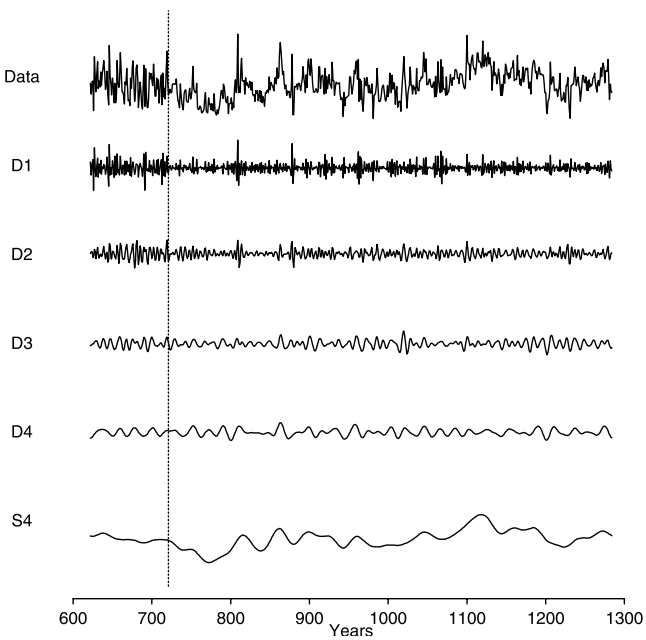

Figure 7. Multiresolution analysis of the Nile River minimum water levels using the D(4) wavelet filter and maximal overlap DWT. The top plot of the figure is the series itself, while the five time series plotted below it constitute an additive decomposition of the series into components associated with, from top to bottom, variations on scales of 1 year ($\widetilde{\mathcal{D}}_1$), 2 years ($\widetilde{\mathcal{D}}_2$), 4 years ($\widetilde{\mathcal{D}}_3$), 8 years ($\widetilde{\mathcal{D}}_4$), and 16 years or longer ($\widetilde{\mathcal{S}}_4$). The vertical dotted line splits the series into two parts: the first 100 observations (622–721 A.D.) and the remaining 563 observations (722–1284 A.D.).

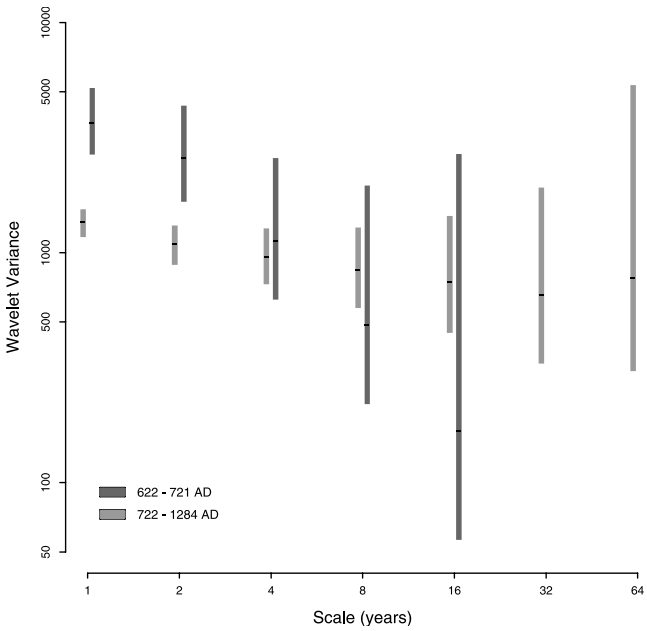

Figure 8. Estimated Haar wavelet variances for the Nile River minimum water levels before and after 721 A.D. The solid bar is the point estimate of the wavelet variance, while the shaded strips are ~95% confidence intervals based upon a chi-square approximation given by *Percival* [1995].

scheme similar to the one used by *Beran* [1994], are displayed in Figure 8. Confidence intervals are equivalent to statistical hypothesis testing at the α level of significance. Thus, since 95% confidence intervals for scales of 1 and 2 years do not overlap, we must reject the null hypothesis that the two wavelet variances are equal at the $\alpha = 0.05$ level of significance. This suggests that the greater variability seen in the first 100 years might be attributable to changes in variance at just these two scales.

[46] For an FD process we have $\nu_j^2 \propto \tau_j^{2d-1}$, so we can estimate d by performing a simple linear regression of $\log \tilde{\nu}_j^2$ on $\log \tau_j$ and by using the estimated slope from the regression $\hat{\beta}$ to form $\hat{d} = \frac{1}{2}\left(\hat{\beta} + 1\right)$. This yields estimates of $\hat{d}$ = 0.38, 0.42, and -0.07 for the whole time series, the last 563 observations, and the first 100 observations, respectively. These compare favorably with Beran's values of 0.40, 0.38, and 0.04, but Figure 8 indicates that the smaller value for $\hat{d}$ in the first 100 years is due to increased variability at scales of 2 years or less. The observed difference in $\tilde{\nu}_j^2$ at longer scales between the first and last portions of the time series is consistent with sampling variability.

[47] Let us now apply the methodology developed in this paper to the Nile River minima from 622 to 1284 A.D. Using all $N = 663$ values in the time series, we computed our test statistic for scales of 1, 2, 4, and 8 years on the basis of 331, 115, 57, and 28 wave let coefficients, respectively. The results from the test, shown in Table 5, confirm an inhomogeneity of variance at scales of 1 and 2 years but fail to reject the null hypothesis of variance homogeneity at scales of 4 and 8 years. Since long memory is a characteristic associated with lower frequencies and variance change is associated with higher frequencies, we infer that there has not been a change in the d but instead a change in the variance of the Nile River series.

[48] Although the time series of Nile River minimum water levels from 622 to 1284 A.D. is the longest continuous record, there are four additional sections of Toussoun's account that span more than 30 years. They are plotted in Figure 9 and span the years 1320–1362 A.D. ($N = 43$), 1364–1433 A.D. ($N = 70$), 1435–1470 A.D. (N = 36), and 1839–1921 A.D. (N = 83). We propose to analyze these four time series in order to show that our methodology is reasonable for hydrological records of lengths more commonly found in practice. Exploratory data analysis was performed via computing the sample auto correlation function (ACF)

$$\hat{\rho}_{Y,\tau} = \frac{\sum_{t=0}^{N-1-\tau} Y_t Y_{t+\tau}}{\sum_{t=0}^{N-1} Y_t^2}, \quad \tau = 0, 1, \ldots, N-1 \tag{43}$$

for each time series. Approximate 95% confidence intervals are provided (dashed lines in Figure 9) under the assumption of Gaussian white noise and are given by

$$\pm \frac{2\sqrt{N-\tau}}{N} \tag{44}$$

[*Fuller*, 1996]. The ACFs for all four time series fall outside these limits at several lags. A portmanteau test of the ACF for no serial correlation [*Ljung and Box*, 1978] was also performed on the four time series and was rejected at the $\alpha = 0.05$ level of significance all four times. Testing homogeneity of variance using a procedure that assumes uncorrelated Gaussian random variables is not appropriate here. Since the lengths of these series are relatively short, we are not assuming they are realizations of FD processes, but rather, they are reasonably smooth across frequency intervals corresponding to the DWT. Thus we may assume that the DWT coefficients are approximately uncorrelated and are able to apply our wavelet-based test for homogeneity of variance on the wavelet coefficients.

[49] A Haar MRA was performed on each of the four series in Figure 9. The only series that indicated a potential change in variance was 1435–1470 A.D., where increased activity was observed in the DWT coefficients before 1450. The monotonic trend in 1839–1921 A.D. is removed by the fact that Daubechies compactly supported wavelets eliminate polynomial functions of order $p = L/2$ from all levels of wavelet coefficients [*Daubechies*, 1992]. Thus the Haar or D(2) wavelet eliminates linear trends while the D(4) wavelet removes quadratic trends and so on. An analysis of the wavelet variance, as in Figure 8, was not performed for these series. Instead, we move directly to computing the test statistic D for the first two scales of the Haar DWT for each series. Table 6 gives the tabulated values for D, and when $\sqrt{N/2}D$ is compared to the Monte Carlo critical values in Table 2, only the first scale of the time series spanning 1435–1470 A.D. rejects the null hypothesis of homogeneity of variance and thus confirms what the MRA identified previously. These four series indicate that our wavelet-based methodology for testing homogeneity of variance in

Table 5. Results of Testing the Nile River Water Levels for Homogeneity of Variance (622–1284 A.D.) using the Haar Wavelet Filter With Monte Carlo Critical Values[a]

Scale	D	10% Critical Level	5% Critical Level	1% Critical Level
1	**0.1559**	0.0945	0.1051	0.1262
2	**0.1754**	0.1320	0.1469	0.1765
4	0.1000	0.1855	0.2068	0.2474
8	0.2313	0.2572	0.2864	0.3436

[a] As shown in the table, the test statistic at scale 1 is significant at the 1% level, and the test statistic at scale 2 is significant at the 5% level. Boldfaced values are statistically significant at the 0.05 (5%) level.

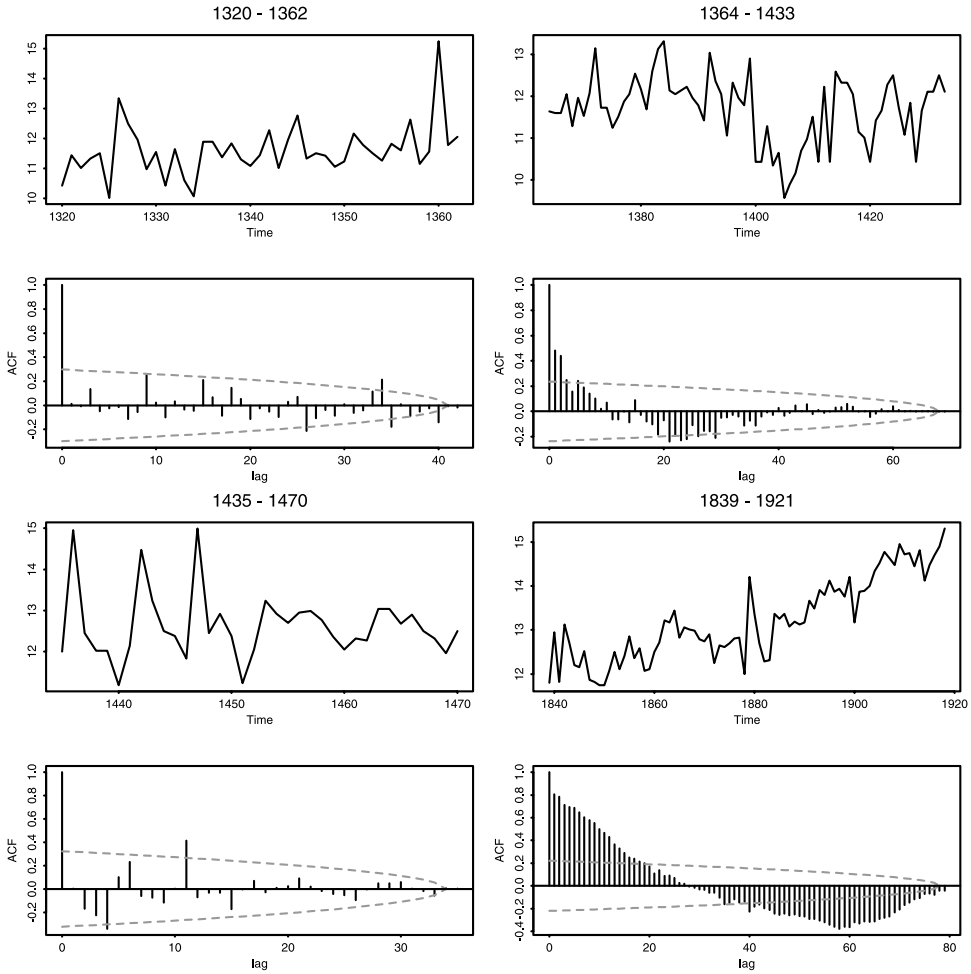

Figure 9. Nile River minimum water level series after 1284 A.D., each of length $N > 30$. Autocorrelation functions (ACFs) with ~95% confidence intervals indicate that these time series are not realizations of Gaussian white noise.

a time series with unspecified correlation structure is valid when the length of the series is <100, a situation commonly encountered when analyzing yearly hydrological records.

5. Locating the Change in Variance

5.1. Auxiliary Test

[50] We shift our attention to determining the location of a variance change in the original time series. A naive choice of location can be based on the test statistic D, i.e., on the location of the DWT coefficient at which the cumulative sum of squares at level j achieves its maximum. Since the DWT coefficient $W_{j,t}$ is a linear combination of observations from the original series as illustrated in equations (16) and (17), this procedure will yield a range of times trapping the variance change. The downsampling inherent in the DWT, however, causes a loss of resolution in time at each scale. Thus we propose to use the nondecimated coefficients $\widetilde{W}_{j,t}$ to determine more accurately the location of a variance change after detection by the DWT.

Table 6. Test Statistic D for the Nile River Minimum Water Levels in Figure 9 Using the Haar Wavelet Filter[a]

Scale	1320–1362 A.D.	1364–1433 A.D.	1435–1470 A.D.	1839–1921 A.D.
1	0.2888	0.1426	0.5790	0.2743
2	0.5830	0.2895	0.4873	0.2793

[a] When $\sqrt{N/2}D$ is compared to the Monte Carlo critical values in Table 2, only the first scale of the time series spanning 1435–1470 A.D. rejects the null hypothesis of homogeneity of variance.

5.2. Simulation Study

[51] A study was conducted to investigate how well the test statistic D, now using the $\widetilde{W}_{j,t}$ coefficients, locates a single variance change in a series with long memory structure. To do this, we implemented a setup motivated by the Nile River example as follows: (1) Generate a realization of length $N = 663$ from an FD process with specified long memory parameter $d = 0.4$; (2) Add Gaussian random variables with $\sigma^2 = 2.07\delta$ to the first 100 observations of the FD process, where 2.07 is the variance $s_{Y,0}$ of an FD process with $d = 0.4$ and $\sigma_\epsilon^2 = 1$ as given by equation (5a); (3) Compute the $\widetilde{W}_{j,t}$ coefficients for $j = 1, \ldots, 4$ using the Haar, D(4), and LA(8) wavelet filters; and (4) Record the location of the $\widetilde{W}_{j,t}$ coefficient from which the test statistic D attains its value, adjusting for the phase shift of the filter output $\widetilde{W}_{j,t}$ by shifting the location $L_j/2$ units to the left (see *Percival and Mofjeld* [1997] for a discussion of the phase shift properties of the nondecimated DWT).

[52] The above was repeated 10,000 times each for $\delta \in \{0.5, 1, 2, 3\}$. Those estimated locations of the variance changes are displayed in Figure 10 for the first two levels of the wavelet

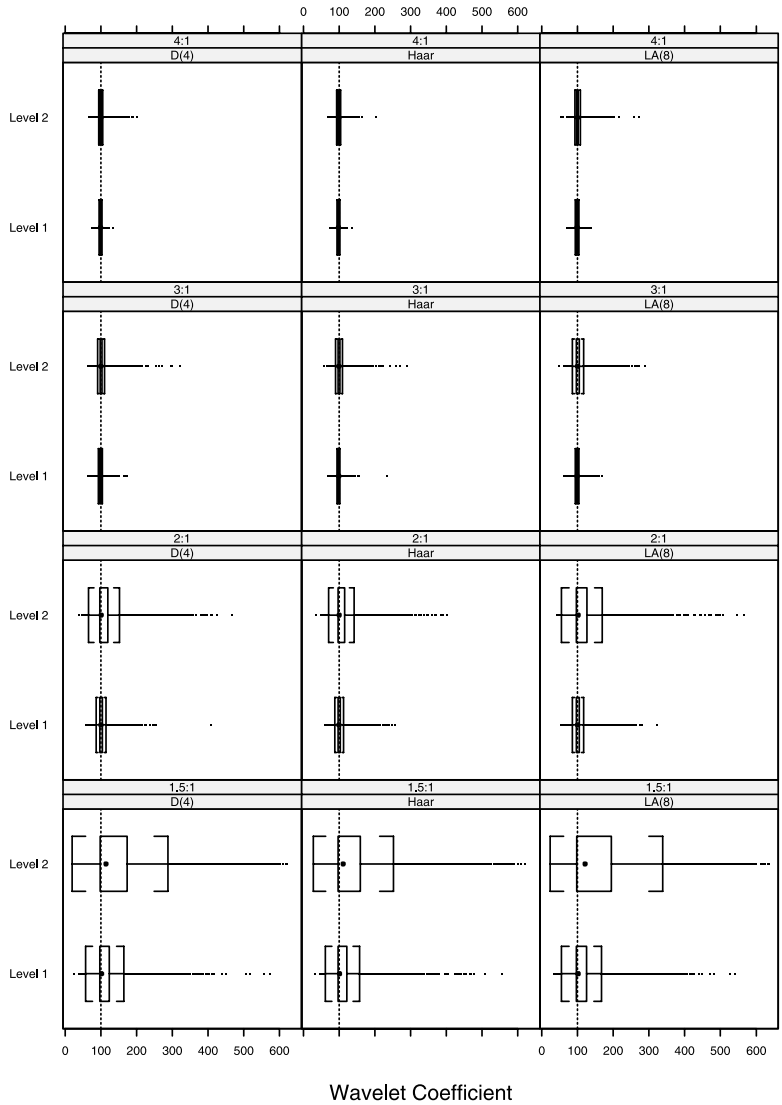

Figure 10. Estimated locations of variance change points for FD processes (N = 663, d = 0.4) using $\widetilde{W}_{j,t}$ with known variance change at t = 100 (dotted line). From bottom to top the variance ratio between the first 100 and remaining observations is 1.5:1, 2:1, 3:1, and 4:1. From left to right the wavelet filters are the D(4), Haar, and LA(8). Under each combination of wavelet filter and variance ratio only the first two levels of the nondecimated DWT are shown (associated with changes of one and two units, respectively).

transform. The columns of Figure 10 correspond to different wavelet filters, the D(4), Haar, and LA(8). The rows correspond to the ratio of the variance between the first 100 observations and those after, ranging from 1.5:1 to 4:1. Each element of Figure 10 contains two box and whisker plots summarizing the distribution of change point locations for the simulation study. The dot is the median, the left and right sides of the box are the lower and upper quartiles, and the "whiskers" are drawn at 1.5 times the quartiles. All observations outside the whiskers are drawn as individual points.

[53] The estimates are roughly centered around the 100th wavelet coefficient for levels j = 1,2 (scales τ_1 and τ_2) with the spread narrowing as the variance ratio increases. There is a very slight difference between wavelet filters, the broader spread being associated with the longer wavelet filters. However, for variance ratios of 2:1 or greater, all three wavelets appear to perform equally well. The estimates from level j = 1 (scale τ_1) have a median value closer to the truth with much less spread at every combination of variance ratio and wavelet filter when compared to the second level. We therefore recommend using the unit-scale estimate when trying to locate an sudden change of variance in a time series.

5.3. Application to the Nile River Water Levels

[54] We apply the above procedure to locate the variance change in the Nile River minimum water levels spanning 622–1284 A.D. Figure 11 displays the normalized cumulative sum of squares as a function of wavelet coefficient for the first two levels. We see a sudden accumulation of variance in the first 100 years and a gradual tapering off of the variance (by construction the series must begin and end at 0). The maximum is actually attained in 720 A.D. for the $\widetilde{W}_{1,t}$ (level 1) coefficients and 722 A.D. for $\widetilde{W}_{2,t}$ (level 2) coefficients. The subsequent smaller peaks occurring in the 9th century are associated with large observations, as seen in the original series, not changes in the variance.

[56] The source document for this series [*Toussoun*, 1925] and studies by *Popper* [1951] and *Balek* [1977] all indicate the construction in 715 A.D. (or soon thereafter) of a "nilometer" in

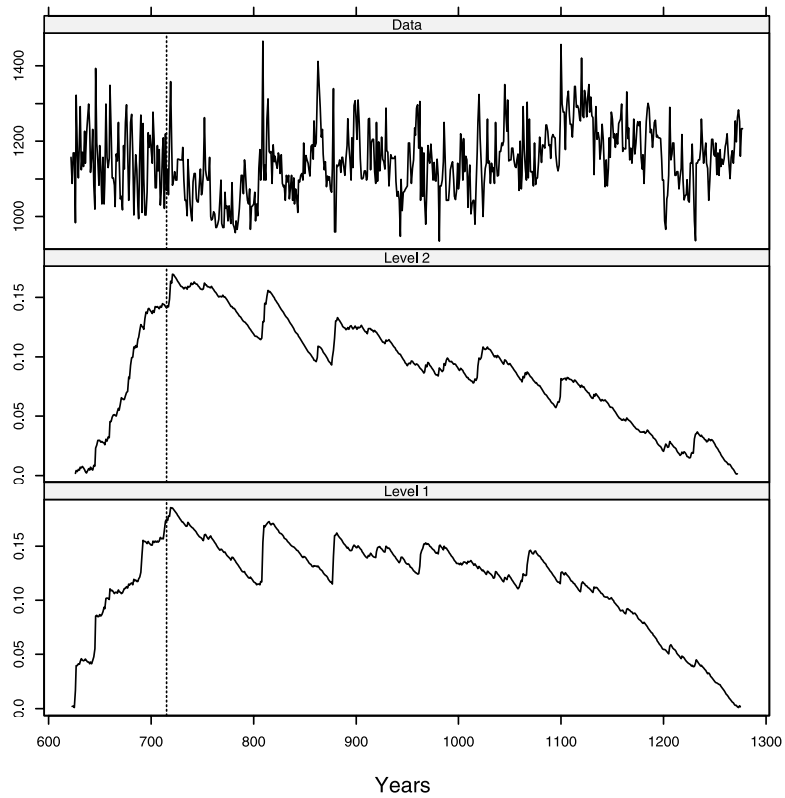

Figure 11. Nile River minimum water levels (upper panel) and normalized cumulative sum of squares using the $\widetilde{W}_{1,t}$ (level 1) and $\widetilde{W}_{2,t}$ (level 2) coefficients based upon the D(4) wavelet filter for the Nile River minimum water levels (lower two panels). The vertical dotted line marks the year 715 AD, after which a nilometer on Roda Island was used to record the water levels.

a mosque on Roda Island in the Nile River. After its construction the yearly minimum water levels up to 1284 A.D. were measured using either this device or a reconstruction of it in 861 A.D. How measurements were made prior to 715 A.D. is unknown, but most likely, devices with less accuracy than the Roda Island nilometer were used. Our estimated change point at 720 or 722 A.D. coincides well with the construction of this new instrument, and it is reasonable that this new nilometer led to a reduction in variability at the very smallest scales.

[56] For the Nile River minimum water level series spanning 1435–1470 A.D. the maximum of the auxiliary test statistic for the $\widetilde{W}_{1,t}$ (level 1) coefficients was attained in 1448 A.D. This provides a nice partition between the more variable beginning of the series and the quiescent portion later on.

6. Discussion

[57] The discrete wavelet transform has been shown to adequately decorrelate time series with long memory structure for the purpose of evaluating a normalized cumulative sum of squares test statistic. It provides a convenient method for detecting and locating inhomogeneities of variance in such time series. The DWT produces a test statistic that can be evaluated under the assumption of white noise, while the nondecimated coefficients $\widetilde{W}_{j,t}$ offer good time domain resolution for locating a variance change. This methodology should be a useful analysis tool applicable to a wide variety of physical processes. See *Whitcher et al.* [2000b] for an extension to multiple variance changes with application to a geophysical time series.

[58] *Ogden and Parzen* [1996] use a test statistic, similar to D, as a solution to a change-point problem for nonparametric regression. Their statistic involves estimating the standard deviation of the squared wavelet coefficients. We avoid this estimation by dividing the cumulative sum by the total sum of squares (see *Whitcher* [1998] for a discussion on the relative merits of these two approaches). Whereas we are looking for changes in the variance, they looked at changes in the mean of a process in order to determine an appropriate level-dependent wavelet threshold. The similarities between the Ogden-Parzen test statistic and D indicate that, while we have discussed D in the context of detecting changes in variance, in fact, D can pick up others kinds of nonstationarities, a fact that must be taken into account before drawing any conclusion when D rejects the null hypothesis of variance homogeneity. Extreme observations (outliers) or sudden changes in the mean of the process may also lead to rejections of the null hypothesis. We recommend performing some exploratory data analysis before applying this methodology.

[59] *Beran and Terrin* [1996] looked at the Nile River minimum water levels (622–1284 A.D.) and used a test statistic to argue for a change in the long memory parameter in the time series. The results from our analysis, in conjunction with an examination of the historical record, suggest an alternative interpretation. There is a decrease in variability at scales of 2 years and less after ~720 A.D., and this decrease is due to a new measurement instrument rather than a change in the long-term characteristics of the Nile River. When analyzing much shorter portions of the Nile River record, our testing procedure is able to eliminate serial correlation and low-order polynomial trends while still being able to test for homogeneity of variance, thus allowing researchers to perform hypothesis

tests on observed time series with potentially complicated trend and correlation structure.

Appendix A: Derivation of Equation (31)

[60] The first step is to separate $\mathcal{H}_j^{(\mathrm{D})}(f)$ into the product of $\mathcal{D}(f)$, $\mathcal{C}(f)$ and $\mathcal{G}(f)$ using an equivalent formula to equation equation (14) for squared gain functions; specifically,

$$\mathcal{H}_j^{(\mathrm{D})}(f) = \mathcal{H}_1^{(\mathrm{D})}\left(2^{j-1}f\right)\mathcal{G}_{j-1}^{(\mathrm{D})}(f) = \mathcal{D}^{\frac{L}{2}}\left(2^{j-1}f\right)\mathcal{C}\left(2^{j-1}f\right)\mathcal{G}_{j-1}^{(\mathrm{D})}(f), \tag{A1}$$

where

$$\mathcal{G}_1^{(\mathrm{D})}(f) = 2\cos^L(\pi f)\sum_{l=0}^{L/2-1}\binom{L/2-1+l}{l}\sin^{2l}(\pi f) \tag{A2}$$

is the squared gain function for the Daubechies scaling coefficients. Using the trigonometric identity, $\sin^2(2f) = 4\sin^2(f)\cos^2(f)$, we may re-express the first term of equation (A1) via

$$\mathcal{D}\left(2^{j-1}\right) = \mathcal{D}(f)\prod_{k=0}^{j-2} 4\cos^2\left(\pi 2^k f\right). \tag{A3}$$

Since we are downsampling by 2 at each level of the DWT, the SDF for the DWT coefficients $W_{j,0},\ldots,W_{j,N/2^j}$ of an FD process Y_t is given by equation (31), i.e.,

$$S_{Y,j}^{(\mathrm{D})}(f) = \frac{1}{2^j}\sum_{k=0}^{2^j-1}\mathcal{H}_j^{(\mathrm{D})}\left(\frac{f+k}{2^j}\right)S_Y\left(\frac{f+k}{2^j}\right), \tag{A4}$$

where

$$\mathcal{H}_j^{(\mathrm{D})}(f) = \mathcal{D}^{\frac{L}{2}}(f)\left[\prod_{k=0}^{j-2} 4\cos^2\left(2^k\pi f\right)\right]^{\frac{L}{2}}\times \quad \mathcal{C}\left(2^{j-1}f\right)\mathcal{G}_{j-1}^{(\mathrm{D})}(f). \tag{A5}$$

Appendix B: Derivation of Equation (40)

[61] Let us assume that the wavelet coefficients are not affected by boundary conditions. We then can write

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l}Y_{2^j(t+1)-1-l} \tag{B1}$$

with a similar expression for $W_{j',t'}$. The covariance between $W_{j,t}$ and $W_{j',t'}$ is given by

$$\mathrm{Cov}\left\{W_{j,t}, W_{j',t'}\right\} = \sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} h_{j,l}h_{j',l'}s_{Y,2^j(t+1)-l-2^{j'}(t'+1)+l'} \tag{B2}$$

and follows from the covariance function being applied to linear combinations of random variables with

$$\mathrm{Cov}\left\{Y_{2^j(t+1)-1-l}, Y_{2^{j'}(t'+1)-1-l'}\right\} = s_{Y,2^j(t+1)-l-2^{j'}(t'+1)+l'}. \tag{B3}$$

Consider $j = j'$ and $t' = t + \tau$. Equation (B3) then becomes

$$\mathrm{Cov}\left\{W_{j,t}, W_{j',t'}\right\} = \sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} h_{j,l}h_{j,l'}s_{Y,2^j\tau+l-l'}. \tag{B4}$$

We can regard the right-hand side of equation (B4) as summing all elements in a symmetric $L_j \times L_j$ matrix, whose (l, l')th element is given by $h_{j,l}h_{j,l'}s_{Y,2^j\tau+l-l'}$ By summing across offset diagonals in this symmetric matrix, we obtain equation (40).

Appendix C: Confidence Intervals for Equation (42)

[63] Here we derive an approximate confidence interval for the estimated wavelet variance using an equivalent degrees of freedom argument [*Priestley*, 1981, p. 466]. Instead of a confidence interval based on the Gaussian distribution for the estimated wavelet variance, we instead claim that

$$\frac{\xi\tilde{\nu}_j^2}{\tilde{\nu}_j^2} \overset{\mathrm{d}}{=} \chi_\xi^2, \tag{C1}$$

where χ_ξ^2 is a chi-squared distribution with ξ degrees of freedom and where "$\overset{d}{=}$" denotes equal distribution. Using the properties of the mean and variance of the χ^2 distribution and appealing to the large-sample approximations to the mean and variance of $\tilde{\nu}_j^2$, we have the following relation

$$\xi = \frac{\left(N - L_j + 1\right)\nu_j^4}{A_j}, \tag{C2}$$

where $A_j = \int_{-1/2}^{1/2} S_{W_j}^2(f)df$ [*Percival*, 1995]. The equivalent degrees of freedom for the chi-squared distribution may be estimated via

$$\hat{\xi} = \frac{\widetilde{N}_j\tilde{\nu}_j^4}{\widehat{A}_j}, \tag{C3}$$

where $\widehat{A}_j$ uses the periodogram to estimate the true SDF. The confidence interval is no longer symmetric under this equivalent degrees of freedom argument because of the χ^2 distribution. Let $q_{\xi,1-\frac{\alpha}{2}}$ and $q_{\xi,\frac{\alpha}{2}}$ be the lower and upper $\alpha/2$ quantiles of the χ^2 distribution with ξ degrees of freedom, i.e.,

$$\mathbf{P}\left[\chi_\xi^2 \le q_{\xi,1-\frac{\alpha}{2}}\right] = \mathbf{P}\left[\chi_\xi^2 \ge q_{\xi,1-\frac{\alpha}{2}}\right] = \frac{\alpha}{2}. \tag{C4}$$

With these quantiles so defined, we have approximately

$$\mathbf{P}\left[q_{\hat{\xi},1-\frac{\alpha}{2}} < \frac{\hat{\xi}\tilde{\nu}_j^2}{\nu_j^2} < q_{\hat{\xi},\frac{\alpha}{2}}\right] = 1-\alpha, \tag{C5}$$

and therefore we may define an approximate $(1-\alpha)$ confidence interval for ν_j^2 via

$$\left[\frac{\hat{\xi}\tilde{\nu}_j^2}{q_{\hat{\xi},1-\frac{\alpha}{2}}}, \frac{\hat{\xi}\tilde{\nu}_j^2}{q_{\hat{\xi},\frac{\alpha}{2}}}\right], \tag{C6}$$

where $\hat{\xi}$ is estimated via equation (C3) in practice.

[63] **Acknowledgments.** The authors would like to thank the Editor, Deputy Editor, and two referees for suggestions that greatly improved this manuscript. This research was instigated while the first two authors were at the Department of Statistics, University of Washington. Partial support for B. Whitcher was graciously provided by the National Science Foundation

under grants DMS98-15344 and DMS93-12686 for the Geophysical Statistics Project and its research.

References

Abraham, B., and W. W. S. Wei, Inferences about the parameters of a time series model with changing variance, *Metrika*, *31*, 183–194, 1984.

Antoniadis, A., and G. Oppenheim (Eds.), *Wavelets and Statistics*, vol. 103, *Lecture Notes in Statistics*, Springer-Verlag, New York, 1995.

Balek, J., *Hydrology and Water Resources in Tropical Africa*, vol. 8, *Developments in Water Science*, Elsevier Sci., New York, 1977.

Bartlett, M. S., *An Introduction to Stochastic Processes With Special Reference to Methods and Applications*, 1st ed., Cambridge Univ. Press, New York, 1955.

Beran, J., *Statistics for Long-Memory Processes*, vol. 61, *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, 1994.

Beran, J., and N. Terrin, Testing for a change of the long-memory parameter, *Biometrika*, *83*, 627–638, 1996.

Billingsley, P., *Convergence of Probability Measures*, John Wiley, New York, 1968.

Box, G. E. P., and G. M. Jenkins, *Time Series Analysis: Forecasting and Control, Time Series Analysis and Digital Processing*, 2nd ed., Holden-Day, Boca Raton, Fla., 1976.

Brooks, C. E. P., *Climate Through the Ages*, 2nd ed., McGraw-Hill, New York, 1949.

Brown, R. L., J. Durbin, and J. M. Evans, Techniques for testing the constancy of regression relationships over time, *J. R. Stat. Soc., Ser. B*, *37*, 149–163, 1975.

Bruce, A., and H.-Y. Gao, *Applied Wavelet Analysis With S-PLUS*, Springer-Verlag, New York, 1996.

Chui, C. K., *Wavelets: A Mathematical Tool for Signal Analysis*, vol. 1, *SIAM Monographs on Mathematical Modeling and Computation*, Soc. for Ind. and Appl. Math., Philadelphia, Pa., 1997.

Coifman, R. R., and D. Donoho, Time-invariant wavelet de-noising, in *Wavelets and Statistics*, vol. 103, *Lecture Notes in Statistics*, edited by A. Antoniadis and G. Oppenheim, pp. 125–150, Springer-Verlag, New York, 1995.

Daubechies, I., *Ten Lectures on Wavelets*, vol. 61, *CBMS-NSF Regional Conference Series in Applied Mathematics*, Soc. for Ind. and Appl. Math., Philadelphia, Pa., 1992.

Davis, W. W., Robust methods for detection of shifts of the innovation variance of a time series, *Technometrics*, *21*, 313–320, 1979.

Eltahir, E. A. B., and G. Wang, Nilometers, El Niño, and climate variability, *Geophys. Res. Lett.*, *26*, 489–492, 1999.

Fuller, W. A., *Introduction to Statistical Time Series*, 2nd ed., Wiley-Interscience, New York, 1996.

Graf, H.-P., Long-range correlations and estimation of the self-similarity parameter, Ph.D. thesis, Eidg. Tech. Hochsch., Zürich, 1983.

Granger, C. W. J., and R. Joyeux, An introduction to long-memory time series models and fractional differencing, *J. Time Ser. Anal.*, *1*, 15–29, 1980.

Hosking, J. R. M., Fractional differencing, *Biometrika*, *68*, 165–176, 1981.

Hsu, D.-A., Tests for variance shift at an unknown time point, *Appl. Stat.*, *26*, 279–284, 1977.

Hsu, D.-A., Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis, *J. Am. Stat. Assoc.*, *74*, 31–40, 1979.

Hurst, H. E., Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.*, *116*, 770–779, 1951.

Inclán, C., and G. C. Tiao, Use of cumulative sums of squares for retrospective detection of changes of variance, *J. Am. Stat. Assoc.*, *89*, 913–923, 1994.

Jarvis, C. S., Flood-stage records of the river Nile, *Trans. Am. Soc. Civ. Eng.*, *101*, 1012–1071, 1936.

Johnson, R. A., and M. Bagshaw, The effect of serial correlation on the performance of CUSUM tests, *Technometrics*, *16*, 103–112, 1974.

Kumar, P., Role of coherent structures in the stochastic-dynamic variability of precipitation, *J. Geophys. Res.*, *101*, 26,393–26,404, 1996.

Kumar, P., and E. Foufoula-Georgiou, Wavelet analysis for geophysical applications, *Rev. Geophys.*, *35*, 385–412, 1997.

Ljung, G. M., and G. E. P. Box, On a measure of lack of fit in time series models, *Biometrika*, *65*, 297–304, 1978.

MacNeill, I. B., S. M. Tang, and V. K. Jandhyala, A search of the source of the Nile's change-points, *Environmetrics*, *2*, 341–375, 1991.

Mallat, S., A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intel.*, *11*, 674–693, 1989.

Mallat, S., *A Wavelet Tour of Signal Processing*, 1st ed., Academic, San Diego, Calif., 1998.

Mandelbrot, B. B., and J. W. van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.*, *10*, 422–437, 1968.

Mandelbrot, B. B., and J. R. Wallis, Some long-run properties of geophysical records, *Water Resour. Res.*, *5*, 321–340, 1969.

McCoy, E. J., and A. T. Walden, Wavelet analysis and synthesis of stationary long-memory processes, *J. Comput. Graphical Stat.*, *5*, 26–56, 1996.

Mohr, D. L., Modeling data as a fractional gaussian noise, Ph.D. thesis, Princeton Univ., Princeton, N.J., 1981.

Molz, F. J., H. H. Liu, and J. Szulga, Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions, *Water Resour. Res.*, *33*, 2273–2286, 1997.

Nason, G. P., and B. W. Silverman, The stationary wavelet transform and some statistical applications, in *Wavelets and Statistics*, vol. 103, *Lecture Notes in Statistics*, edited by A. Antoniadis and G. Oppenheim, pp. 281–300, 1995.

Nuri, W. A., and L. J. Herbst, Fourier methods in the study of variance fluctuations in time series analysis, *Technometrics*, *11*, 103–113, 1969.

Ogden, R. T., and E. Parzen, Change-point approach to data analytic wavelet thresholding, *Stat. Comput.*, *6*, 93–99, 1996.

Percival, D. B., On estimation of the wavelet variance, *Biometrika*, *82*, 619–631, 1995.

Percival, D. B., and P. Guttorp, Long-memory processes, the Allan variance and wavelets, in *Wavelets in Geophysics*, vol. 4, *Wavelet Analysis and Its Applications*, edited by E. Foufoula-Georgiou and P. Kumar, pp. 325–344, Academic, San Diego, Calif., 1994.

Percival, D. B., and H. O. Mofjeld, Analysis of subtidal coastal sea level fluctuations using wavelets, *J. Am. Stat. Assoc.*, *92*, 868–880, 1997.

Perreault, L., J. Bernier, B. Bobée, and E. Parent, Bayesian change-point analysis in hydrometeorological time series, part 1, The normal model revisited, *J. Hydrol.*, *235*, 221–241, 2000.

Popper, W., *The Cairo Nilometer*, vol. 12, *Publications in Semitic Philology*, Univ. of Calif. Press, Berkeley, Calif., 1951.

Priestley, M. B., *Spectral Analysis and Time Series*, Academic, San Diego, Calif., 1981.

Shensa, M. J., The discrete wavelet transform: Wedding the à trous and Mallat algorithms, *IEEE Trans. Signal Process.*, *40*, 2464–2482, 1992.

Stephens, M. A., EDF statistics for goodness of fit and some comparisons, *J. Am. Stat. Assoc.*, *69*, 730–737, 1974.

Tang, S. M., and I. B. MacNeill, The effect of serial correlation on tests for parameter change at unknown time, *Ann. Stat.*, *21*, 552–575, 1991.

Tewfik, A. H., and M. Kim, Correlation structure of the discrete wavelet coefficients of fractional Brownian motion, *IEEE Trans. Inf. Theory*, *38*, 904–909, 1992.

Toussoun, O., Mémoire sur l'histoire du Nil, in *Mémoires à l'Institut d'Egypte*, vol. 18, pp. 366–404, 1925.

Tsay, R. S., Outliers, level shifts, and variance changes in time series, *J. Forecasting*, *7*, 1–20, 1988.

Wang, Y., Jump and sharp cusp detection by wavelets, *Biometrika*, *82*, 385–397, 1995.

Whitcher, B., Assessing nonstationary time series using wavelets, Ph.D. thesis, Univ. of Washington, Seattle, Wash., 1998.

Whitcher, B., P. Guttorp, and D. B. Percival, Wavelet analysis of covariance with application to atmospheric time series, *J. Geophys. Res.*, *105*, 14,941–14,962, 2000a.

Whitcher, B., P. Guttorp, and D. B. Percival, Multiscale detection and location of multiple variance changes in the presence of long memory, *J. Stat. Comput. Simul.*, *68*, 65–88, 2000b.

Wichern, D. W., R. B. Miller, and D.-A. Hsu, Changes of variance in first-order autoregressive time series models—With an application, *Appl. Stat.*, *25*, 248–256, 1976.

Wornell, G. W., *Signal Processing with Fractals: A Wavelet Based Approach*, Prentice-Hall, Old Tappan, N.J., 1996.

S. D. Byers, AT&T Labs Research, 180 Park Ave, P.O. Box 971, Florham Park, NJ 07932-0971, USA. (byers@research.att.com).

P. Guttorp, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322, USA. (peter@stat.washington.edu).

D. B. Percival, Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195-5640, USA. (dbp@apl.washington.edu).

B. Whitcher, Geophysical Statistics Project, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA. (whitcher@ucar.edu).